The structural basis for pyrophosphatase catalysis

Pirkko Heikinheimo[1,2], Jukka Lehtonen[1,2], Alexander Baykov[3], Reijo Lahti[2], Barry S Cooperman[4]* and Adrian Goldman[1,2]*

Background: Soluble inorganic pyrophosphatase (PPase), an essential enzyme central to phosphorus metabolism, catalyzes the hydrolysis of the phosphoanhydride bond in inorganic pyrophosphate. Catalysis requires divalent metal ions which affect the apparent pK_as of the essential general acid and base on the enzyme, and the pK_a of the substrate. Three to five metal ions are required for maximal activity, depending on pH and enzyme source. A detailed understanding of catalysis would aid both in understanding the nature of biological mechanisms of phosphoryl transfer, and in understanding the role of divalent cations. Without a high-resolution complex structure such a model has previously been unobtainable.

Results: We report the first two high-resolution structures of yeast PPase, at 2.2 and 2.0 Å resolution with R factors of around 17 %. One structure contains the two activating metal ions; the other, the product $(MnP_i)_2$ as well. The latter structure shows an extensive network of hydrogen bond and metal ion interactions that account for virtually every lone pair on the product phosphates. It also contains a water molecule/hydroxide ion bridging two metal ions and, uniquely, a phosphate bound to four Mn^{2+} ions.

Conclusions: Our structure-based model of the PPase mechanism posits that the nucleophile is the hydroxide ion mentioned above. This aspect of the mechanism is formally analogous to the 'two-metal ion' mechanism of alkaline phosphatase, exonucleases and polymerases. A third metal ion coordinates another water molecule that is probably the required general acid. Extensive Lewis acid coordination and hydrogen bonds provide charge shielding of the electrophile and lower the pK_a of the leaving group. This 'three-metal ion' mechanism is in detail different from that of other phosphoryl transfer enzymes, presumably reflecting how ancient the reaction is.

Addresses: [1]Turku Centre for Biotechnology, PO Box 123, FIN-20521 Turku, Finland, [2]Department of Biochemistry, University of Turku, FIN-20014 Turku, Finland, [3]A.N. Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119899, Russia and [4]Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA.

*Corresponding authors.
E-mail: cooprman@pobox.upenn.edu
E-mail: Goldman@btk.utu.fi

Key words: mechanism, phosphoanhydride hydrolysis, phosphoryl transfer, pyrophosphatase, refinement, structure

Received: 9 August 1996
Revisions requested: 30 August 1996
Revisions received: 20 September 1996
Accepted: 15 October 1996

Structure 15 December 1996, 4:1491–1508

Introduction

Soluble inorganic pyrophosphatase (PPase; E.C. 3.6.1.1), an ubiquitous enzyme, is essential for cell growth [1,2]. It provides a thermodynamic pull for many biosynthetic reactions which utilize nucleotide triphosphates, including nucleic acid, polysaccharide and protein biosynthesis, by catalyzing the essentially irreversible hydrolysis of the product inorganic pyrophosphate (PP_i) to two phosphates (P_i) [3]. Although phosphoryl transfer enzymes, of which PPase is one, form one of the largest classes of enzymes in nature, their mechanisms are still not completely understood [4]. The comparative simplicity of PP_i as a substrate, and the importance of PPase in the control of overall phosphorus metabolism have prompted our detailed studies of this enzyme. We report here two structures of the PPase from the budding yeast *Saccharomyces cerevisiae* (Y-PPase) that, considered together with the results of functional studies of the wild-type and variant enzymes, permit us to formulate a substantially complete structure-based model for the PPase mechanism.

The first PPase structure appeared in 1981 [5] at 3.0 Å resolution (Protein Data Bank [PDB] code 1pyp [6]). Although an unrefined apostructure, it showed the fold of the enzyme and the active site cleft, demonstrating that the active site is large and contains many polar residues. Later, a product-containing structure of Y-PPase was reported [7], but the coordinates for this structure were never released. As subsequent work by the Harutyunyan group [8] and this work demonstrate, the earlier structure contains errors in the description of product binding. So far, two prokaryotic PPase structures are known at high resolution: the *Escherichia coli* PPase (E-PPase) apoenzyme structure has been solved at 2.2 Å resolution [9–11] (PDB codes 2eip, 1faj, 1ipw, 1igp) and *Thermus thermophilus* PPase (T-PPase) at 2.0 Å resolution [12] (PDB code 2prd). None of these structures contain substrate, but Mg^{2+} was soaked into the E-PPase crystals (Mg·E-PPase) (1ipw) [10]; the T-PPase structure has one sulphate ion mimicking the binding of the phosphate ion and a report of one Mn^{2+} ion in the *E. coli* PPase active site has also appeared [13]. We

report here the first two high resolution Y-PPase structures. The first structure (Mn_2·Y-PPase) contains both the metal ions that are required for the activation of the enzyme and that are present before substrate binds. The second structure is a product complex, Mn_2·Y-PPase·$(MnP_i)_2$, having, in addition to the two activating manganese ions, two phosphates and two additional manganese ions.

PPases hydrolyze PP_i without a covalent enzyme bound intermediate [14]. The reaction involves at least three divalent metal ions in the active site; the relative activity with different cations is Mg^{2+}>Zn^{2+}>Co^{2+}≈Mn^{2+}>>Cd^{2+} [15,16]. Measured at pH 7, at or near the pH optimum, the k_{cat} for the Mn^{2+}–enzyme is about 10% of that for the Mg^{2+}–enzyme, while the microscopic rate constant for hydrolysis of enzyme-bound PP_i is decreased by a factor of five. Mn^{2+} is thus a good model for the natural activating metal ion, and has some crystallographic advantages (see Materials and methods section). The metal ions both activate the enzyme and are also part of the substrate [15,17]. PPases are diffusion-controlled enzymes [15]; by understanding PPase catalysis in detail, we hope to understand better the use of metal ions in catalysis and the enzyme-catalyzed hydrolysis of more complex substrates. The use of metal ions is very common in phosphoryl transfer enzymes. A 'two metal ion' mechanism has been proposed for other phosphoryl transfer enzymes such as, DNA polymerases [18], retroviral integrases [19], HIV reverse transcriptase [20] and alkaline phosphatase [21]. Some have even called it the 'common' two-metal ion mechanism [22] without, however, saying whether the mechanistic similarity is anything more than formal. The three-dimensional structure of alkaline phosphatase, for instance, shows no evidence of being related to the others by divergent evolution.

The best studied PPases are Y-PPase and E-PPase. The active site is very well conserved in all PPases; for instance 13–14 of the 17 polar active-site residues are fully conserved in the structure-based sequence alignments of Y- and E-PPases [23]. However, the overall sequence identity is only 27% [24] and the oligomeric structure of the two enzymes is different, Y-PPase being a homodimer and E-PPase a homohexamer [25]. The catalytic mechanism of these two enzymes is very similar [16,25,26]. Mutations of conserved active-site residues produce rather similar changes in the activities of Y-PPase and E-PPase, although Y-PPase activity [27] is somewhat more sensitive to mutation than E-PPase activity [28].

The reaction mechanism requires a basic and an acidic group in the active site. The apparent pK_a of the basic group seems to be affected by almost every active-site mutation we have studied. Even conservative mutations (Asp↔Glu or Arg↔Lys) of the conserved active-site residues increase the pK_a of the basic group by 1–3 pH units while having smaller effects on the pH independent values of k_{cat} [28]. Based on this and other evidence, we suggested that the essential basic group for the reaction is not an amino acid sidechain, but an hydroxide ion coordinated to metal ion in the PPase active site [27,28]. The shift in the pK_as would thus be due to changes in the overall electrostatic environment in the active site [16,28]. Our two new Y-PPase structures clearly support this hypothesis. Our catalytic model now identifies an hydroxide bridging two metal ions as the nucleophile attacking PP_i. Further it includes proposals for the identity of the general acid, the roles of specific metal ions and active-site residues in PPase, and the differences between Y- and E-PPase mechanisms of action.

Results and discussion

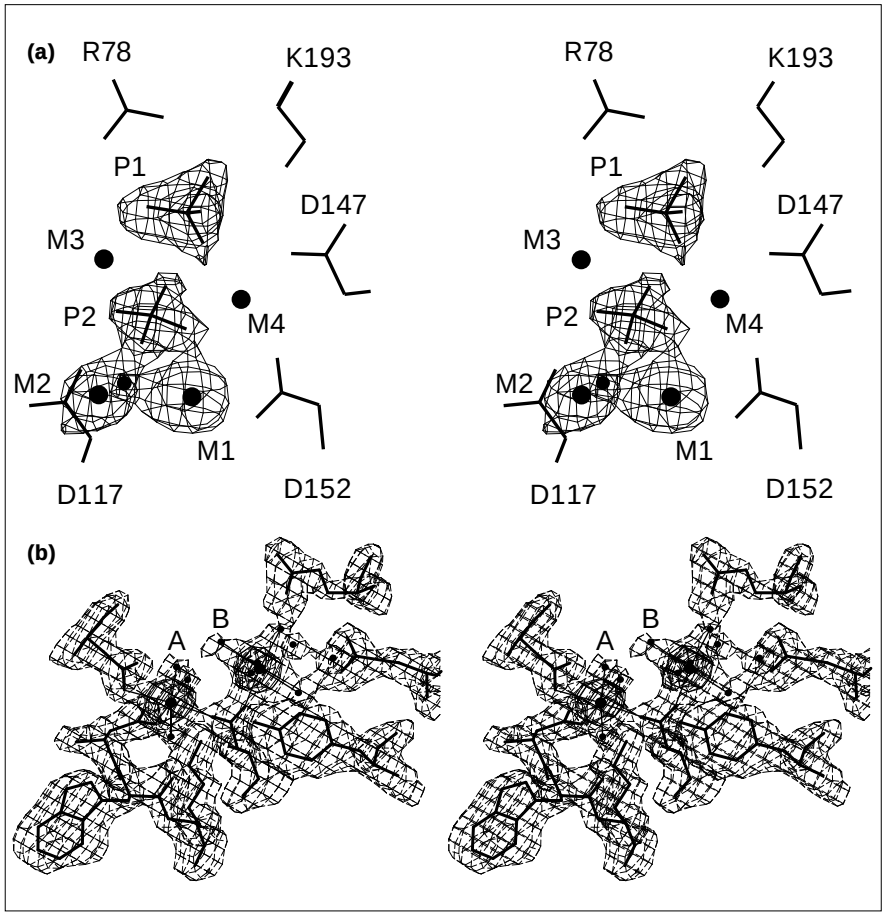
Monomeric structure

We have solved two structures of complexes of Y-PPase, Mn_2·Y-PPase and Mn_2·Y-PPase·$(MnP_i)_2$. The Mn_2·Y-PPase structure was solved by molecular replacement using a partially refined Y-PPase apoenzyme structure (D Voet, personal communication); the Mn_2·Y-PPase·$(MnP_i)_2$ structure was solved by molecular replacement from the Mn_2·Y-PPase structure. No large changes have been made to the overall structure of Y-PPase, compared either with the initial model or with the one in the PDB (1pyp). The structure is now more accurate due to higher resolution and refinement, leading to a somewhat different tracing for one β strand.

The first report of a Y-PPase product complex structure [7] contained errors because they located only three manganese ions in the active site. We have been able to correct the errors by using anomalous difference Fourier maps to position the manganese ions and by refinement to higher resolution. The product complex structure of Harutyunyan *et al.* [8], published at almost exactly the same time as the initial submission of this manuscript, appears to be similar to ours although a detailed comparison has yet to be done. There are some important differences, however, both in structure and mechanistic interpretation, which will be referred to below. Many of the structural differences appear to be due to the quality of the data; our Mn_2·Y-PPase·$(MnP_i)_2$ crystals diffract to better than 1.5 Å (PH, unpublished data), while those of Harutyunyan *et al.* [8] diffract to 2.4 Å at best. Consequently, our structure is at a higher resolution (nominally 2.0 Å versus 2.4 Å) and our effective resolution is in fact the same as our nominal resolution, theirs is almost certainly lower. This can be seen in the refinement statistics and the quality of the electron-density maps (Fig. 1). (Details of the refinement are described in the Materials and methods section.)

The structures compared here present the enzyme in four different states: the *E. coli* apoenzyme (apoE-PPase) [9]; E-PPase with the tightest metal ion bound (Mg·E-PPase) [10]; the fully activated resting enzyme with two metal ions bound (Mn_2·Y-PPase); and the product–enzyme complex (Mn_2·Y-PPase·$(MnP_i)_2$). The two Y-PPase structures have

Figure 1



Electron density for the two Y-PPase structures. (a) Stereo view of the (F_o–F_c) electron density, contoured at 4.5σ in a simulated annealing omit map at 2.0 Å for the Mn_2·Y-PPase·$(MnP_i)_2$ structure. The two phosphates, M1, M2 and Wat1 (see text for definitions) have been omitted to show the quality of the electron density and our ability to define water molecules and the positions of the oxygen atoms on the phosphates. The metal ions and the key active-site residues are labelled. (b) Stereo view of the electron density around metal ions M1 and M2 in the Mn_2·Y-PPase structure. The maps, calculated at 2.2 Å, are ($3F_o$–$2F_c$), shown as dashed lines, and an anomalous difference map, in solid lines; the maps are contoured at 1σ and 5σ, respectively. The current model is shown, with larger circles for the metal ions and smaller circles for the water molecules. The two water molecules in the M1–M2 bridge are labelled A and B and are coordinated to M1 and M2, respectively. (Figures were generated using the program O [29] and OPLOT [M Kjeldgaard, unpublished program].)

no major differences between them (see later) so, unless specified, what follows applies to both.

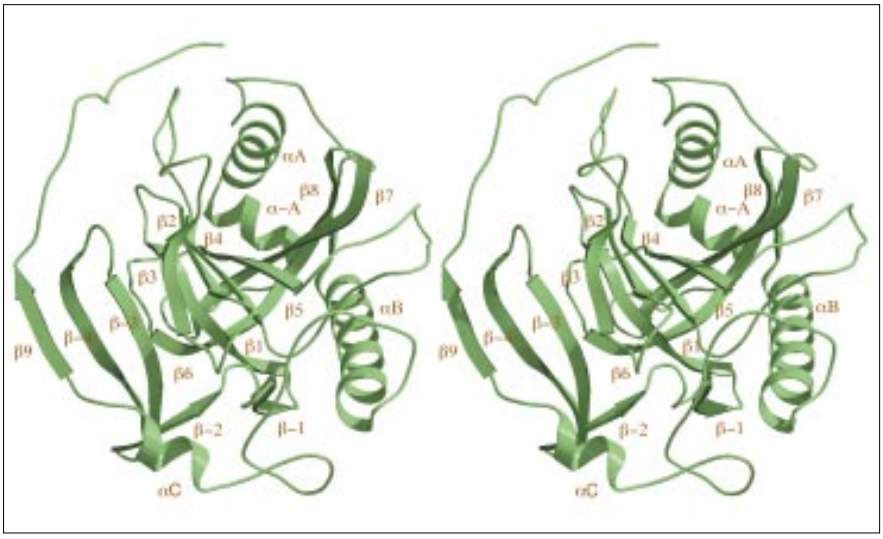
The core of the enzyme is a twisted five stranded β barrel (Fig. 2), formed from strands β1 and β4–β7. The core is conserved in all PPases [9,12] and its top surface forms the base of the active site (Fig. 2). Eukaryotic PPases are longer than prokaryotic ones [25], mainly because of N- and C-terminal extensions that fold on top of the enzyme (Fig. 3). In addition, Y-PPase has another β sheet formed by the N- and C-terminal extension β strands (β-4, β-3, β9) as well as β2 and β3, which form a hairpin loop in E-PPase. Between β3 and β4 there is a long loop (residues 71–91), rather different than in E-PPase, that makes contact with helix αA, contains the active-site residues Arg78 and Tyr89, and makes β-sheet like interactions with the N terminus (Fig. 2). This loop also contains a *cis*-peptide between residues 84 and 85 and is involved in dimer formation (see later). Changing the Arg78–Asp71 interaction affects the thermostability of the enzyme [27]; the Arg78→Lys variant protein is less thermostable than wild type, presumably because the interaction is weakened and so the mobility of the loop is increased. The reverse is true for the Asp71→Glu variant, which is more thermostable than wild type [27], presumably because the carboxylate group of Glu71 is closer to and interacts more strongly with Arg78 than does Asp71 in the wild-type protein.

In addition to the β structure, all PPases have two long α helices, αA and αB. Both conserved helices of Y-PPase align well with the corresponding helices in E-PPase. Preceding helix A, and at about 90° to it, Y-PPase has one shorter α helix (helix-A, residues 171–177). There are also four 3_{10}-helical turns, at residues 165–167, 194–196, 205–207 and 274–277.

Dimer formation

The monomer–monomer contacts in the Y-PPase homodimer are mainly formed by stacking of aromatic rings (Fig. 4). Histidine residues His87 and His87′ are hydrogen bonded to each other in a plane in the centre of the stack. On one side, ~3.5 Å away is Trp52, and on the other side is Trp52′, thus creating a three-layer arrangement reminiscent of base stacking. The long hydrophobic arm of Arg51

Figure 2



Stereo view of the overall fold of Y-PPase, as exemplified by the Mn_2·Y-PPase·$(MnP_i)_2$ structure. The secondary structure elements are labelled; the secondary structure definition is from PROCHECK [67]. The conserved secondary structure (β1–β8 and αA and αB) is named as in E- and T-PPase [9,12]. The secondary structures in Y-PPase that do not occur in E-PPase are marked with decreasing numbering for the N-terminal (β–1, β–2, α–A etc) insertions and with increasing numbering for the C-terminal insertions (β9 and αC). Strands with at least three residues and helices with at least four residues are drawn as arrows and spirals, respectively. In the figure, one residue was added to each end of each secondary structure element for clarity. The active site is located in the space between helix A and the top of the conserved β barrel. (Figure generated using the program MOLSCRIPT [70] and Raster3D [71].)

packs against Trp52 while the buried, most probably charged, headgroup makes all five hydrogen bonds that it can. (Burial of charged groups in proteins is not as rare as is commonly believed [S Helin, PC Kahn and AG, unpublished data].) The Nη1 and Nη2 protons of Arg51 are hydrogen bonded to the mainchain carbonyl oxygens of Asp277′ (across the twofold axis), Glu126 and Gly125, and to a well ordered water molecule, while the Nε proton is hydrogen bonded to the ring electrons of Trp52. In addition to the stacked residues, Trp279 packs perpendicularly against the Trp52′ and Arg51′ residues (and Trp279′ packs against Trp52 and Arg51).

None of the dimer-forming residues are a part of the active site, explaining why none of the active-site mutations affects the oligomeric structure of the enzyme [27], unlike in E-PPase [28] where some active-site mutations cause dissociation of the hexamer into trimers. This dimer interface is likely to be conserved in other eukaryotic PPases. Arg51, Trp52 and Trp279 are conserved in human (EMBL/Genbank entries T61949 and R17778), bovine and all known yeast sequences [25]. His87 is conserved in all yeast sequences but is replaced by a lysine in the bovine retinal and human PPases.

Similarities and differences between the structures

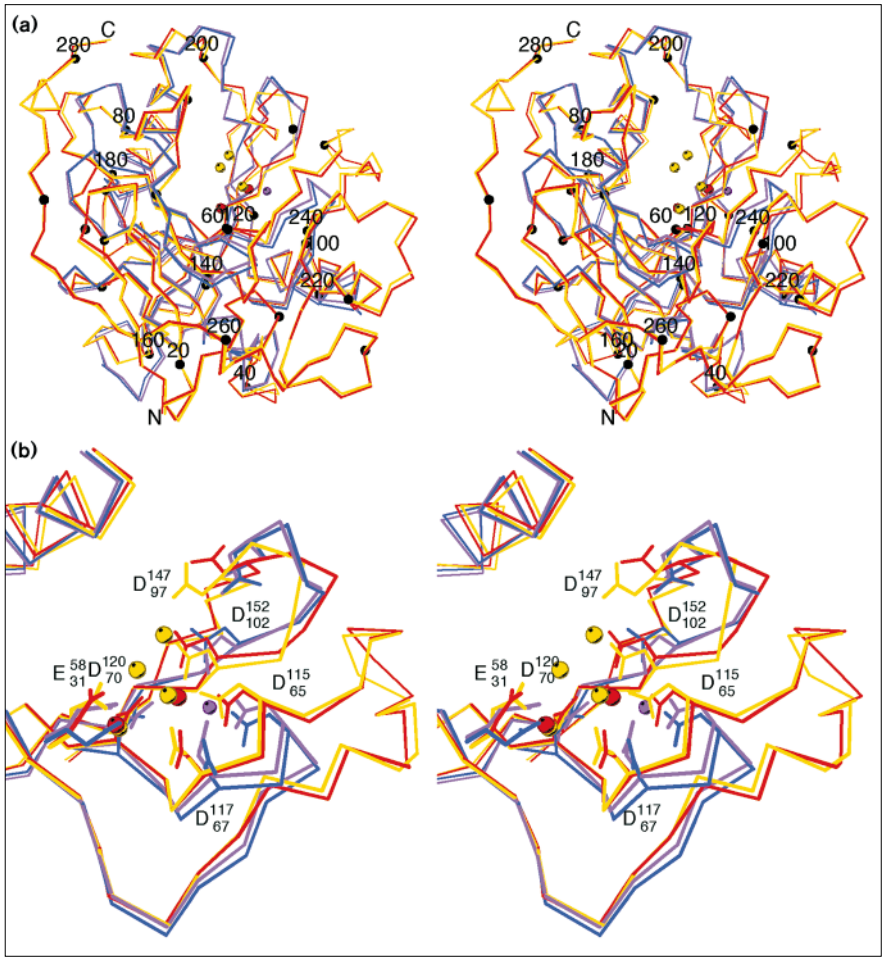
The core structures of E- and Y-PPase are very similar, but Y-PPase has a 27-residue N-terminal and 59-residue C-terminal extension compared with E-PPase [23]. With the initial alignment generated using the program O [29], 154 residues of the Mg·E-PPase structure align with a root mean square (rms) deviation of 1.40 Å to the Mn_2·Y-PPase structure and 154 residues with an rms deviation of 1.45 Å to the Mn_2·Y-PPase·$(MnP_i)_2$ structure. To find out in which areas the structures differed, we refined the superposition with RgbSup (P Fitzgerald, unpublished program), excluding all Cα atoms deviating by more than 1 Å from the superposition. Under these conditions, 87 of the E-PPase residues superimpose on the Mn_2·Y-PPase structure with an rms deviation of 0.66 Å. All the β strands that form the barrel are included in the most similar areas (Fig. 3a). The most different regions are (Y-PPase numbering): Ala28–Asn41, Ile59–Ile90, Glu101–Asn116, Met143–Trp153, Asp159–Asp165 and Pro195–Glu208; the rms deviations per Cα in these regions, after refinement by RgbSup which excludes most of these residues, are 4.06, 2.82, 1.99, 1.94, 4.41 and 2.92 Å, respectively.

The two metal ion binding loops (Glu101–Asn116 and Met143–Trp153; Fig. 3b), among the most different regions, contribute to the active site being much more open in E-PPase than in Y-PPase (see later). The former loop is much shorter in E-PPase (Ser63–Gly66) than in Y-PPase, and the latter, though the same length, has a very different conformation (Fig. 3b). Some, but not all, of the differences in these loops are due to the different metallation states in these structures. This is demonstrated by the fact that the Glu101–Asn116 and Met143–Trp153 regions are the ones that differ most between the two Y-PPase structures (Fig. 3b). The respective Y-PPase/Y-PPase rms deviations per Cα are 1.01 Å (Glu101–Asn116) and 1.27 Å (Met143–Trp153), after refinement by RgbSup, versus an overall Y-PPase/Y-PPase deviation of around 0.5 Å. (Further details can be found within the supplementary material.)

Binding of the activating metal ions

Two activating manganese ions bind at the base of the active site at sites M1 and M2 in the absence of substrate. Mn^{2+} at M1 is coordinated by three aspartic residues, Asp115, Asp120 and Asp152 (Fig. 1b). The remaining three

Figure 3



Stereo view superimpositions of four PPase structures. **(a)** Superimposition of Cα traces of four PPase structures: the Mn_2·Y-PPase·$(MnP_i)_2$ structure is in yellow, the Mn_2·Y-PPase structure is in red, the Mg·E-PPase structure is in purple and the apoE-PPase structure is in blue. Metal ions are presented as spheres with the color of the corresponding structure. Every tenth residue in the Mn_2·Y-PPase·$(MnP_i)_2$ structure is indicated by a black ball and every twentieth residue is labelled. **(b)** Close-up stereo view of the superimposition in Figure 3a, from a slightly rotated view; the colouring is as in Figure 3a. The difference in the position of M1 in the three metal ion containing structures is apparent. The major backbone motion in the loops 101–115 and 144–152 can be seen, as can the motions of the metal-binding residues: Glu58, Asp115, Asp117, Asp120, Asp147 and Asp152. (Figure generated using the program MOLSCRIPT [70].)

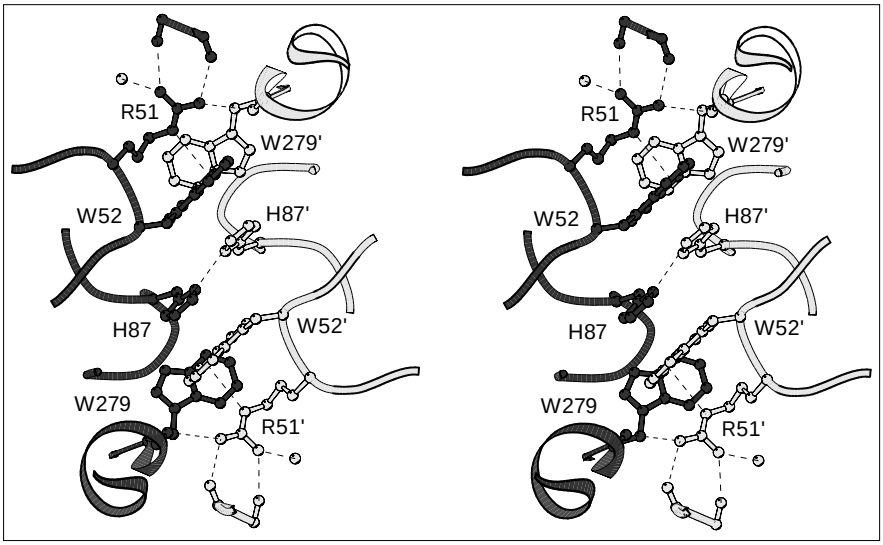
coordination sites are filled by water molecules and the coordination is almost octahedral. Mn^{2+} at M2 is directly coordinated to only one protein residue, Asp120. The rest of its coordination sites are filled by water molecules in the Mn_2·Y-PPase structure; one of the waters is further hydrogen bonded to Glu48.

For Y-PPase, the macroscopic dissociation constant K_{m1} ([Mn][E]/[MnE]) is 10 μM, and K_{m2} ([Mn][MnE]/[Mn2E]) is 40 μM [30], measured at pH 7.2. They thus differ by the statistical factor of four, as expected for dissociation from two sites of equivalent affinity (i.e. M1 and M2 have similar intrinsic metal ion affinities). In contrast, for E-PPase at the same pH, the K_{m2}/K_{m1} ratio is equal to 25 for Mg^{2+} binding [31] and 40 for Mn^{2+} binding (Hyytiä *et al.*, unpublished data) so one site clearly binds more tightly than the other. For E-PPase, the tighter site may be assigned to M1 (Fig. 5) for two reasons. Firstly, M1 is occupied even when site M2 is not, in Mg·E-PPase [10] and in Mn·E-PPase [13]. Secondly, the E-PPase Asp65→Glu and Asp102→Glu variants (Y-PPase Asp115 and Asp152) show weakened binding for the high affinity site but not for the low affinity site (Hyytiä *et al.*, unpublished data). Surprisingly, the Asp70→Glu variant (Y-PPase Asp120) binds metal ions at both sites with the same affinity as wild type (Hyytiä *et al.*, unpublished data), but is almost completely inactive [28]. Whatever change occurs to allow metal ion binding must also prevent productive binding of the substrate.

Product binding

The two P_is in the Mn_2·Y-PPase·$(MnP_i)_2$ structure bind at sites denoted P1 and P2. In the Mn_2·Y-PPase·$(MnP_i)_2$ structure, coordination to P_i in site P2 replaces coordination to one water molecule at both M1 and M2 (Fig. 5) and the distance between these metal ions decreases, from 4.9 Å to 3.7 Å in Mn_2·Y-PPase (Fig. 3b, Tables 1,2). As a consequence, whereas two water molecules form the M1–M2 bridge in Mn_2·Y-PPase (Fig. 1b), a single water molecule, quite important for the hydrolysis mechanism (see below) is able to bridge M1 and M2 in Mn_2·Y-PPase·$(MnP_i)_2$ (Fig. 1a). Harutyunyan *et al.* [8] were unable to see a bridging water molecule between M1 and M2 and their hydrolysis

Figure 4



Stereo view of the dimer–dimer contact in Y-PPase. One monomer is drawn in dark grey and the other in light grey. Hydrogen bonds are drawn as dashed lines. The interactions of the Arg51 headgroup are shown, including the hydrogen bond to the ring of Trp52. (Figure generated using the program MOLSCRIPT [70].)

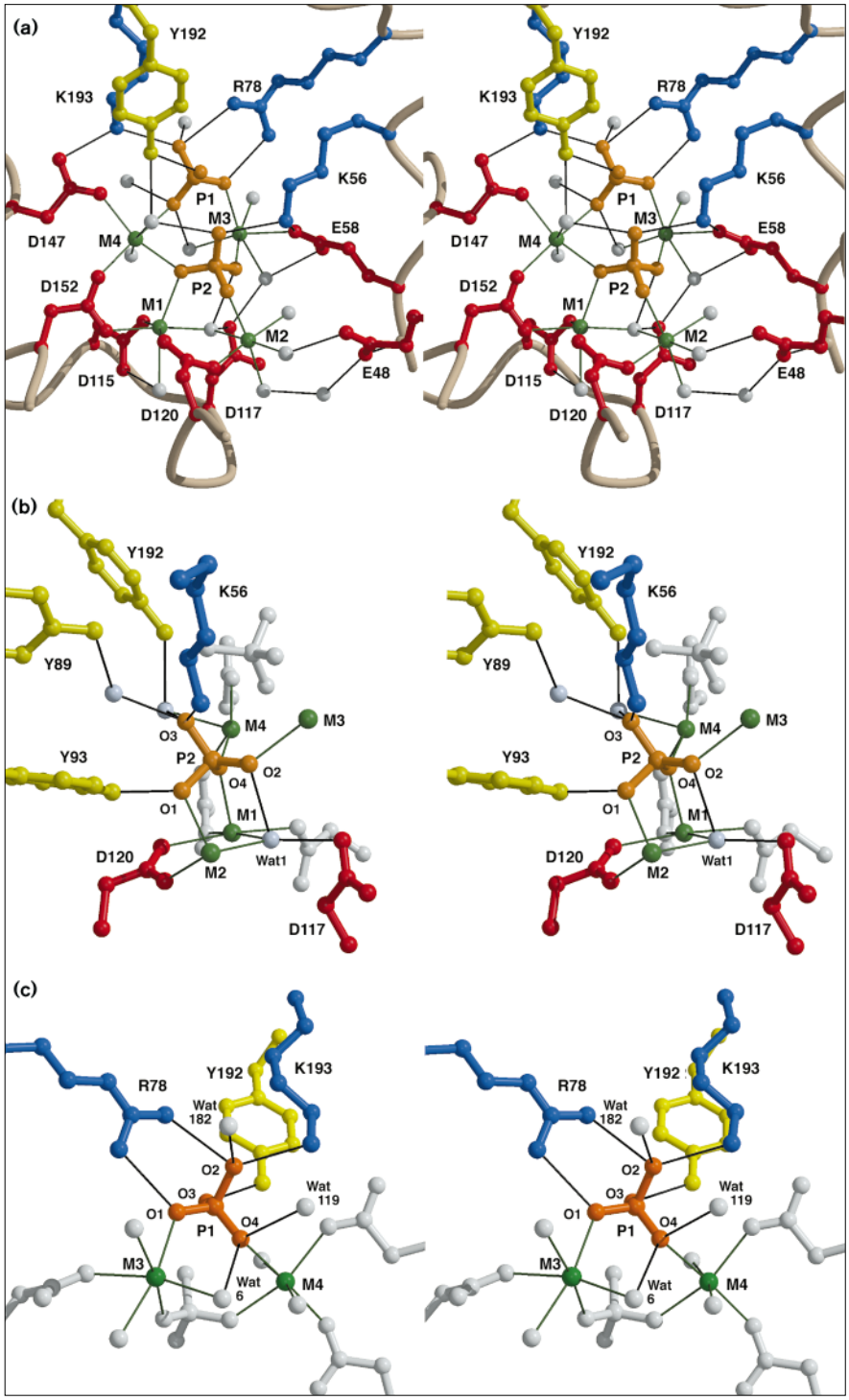
mechanism fails to explain convincingly the need for both M1 and M2. Mn^{2+}–Mn^{2+}, Mn^{2+}–P_i, and P_i–P_i distances are presented in Table 1; these values are both more precise and somewhat different from those presented in Harutyunyan *et al.* [8]. These authors, interpreting a lower resolution structure, had more difficulty than we did in deciding which peaks in the active site were metal ions, which were phosphate ions, where the oxygens on the phosphates point, and what the coordination of the metal ions is.

Mn^{2+} in sites M3 and M4 bind to both P_is and are both coordinated by sidechain carboxylates: M3 to Glu58 and M4 to Asp147 and Asp152. The binding of product induces a change in the active-site conformation due to movement in two loops: Glu101–Asn116 (from the end of β4 to the beginning of β5) and Met143–Trp153 (from the end of β6 to the beginning of β7) (Fig. 3b). The movement of the latter is more pronounced than the movement of the former. The motion positions Asp147 and Asp152 (Fig. 3) such that Asp152 coordinates M1 and M4, while Asp147 coordinates M4 and makes a hydrogen bond with the substrate-binding residue Lys193. Residues Asp102 and Val105–Pro108 of the first segment are also involved in crystallographic contacts in the Mn_2·Y-PPase structure. Movement of this region is clearly due to product binding, as soaking the Mn_2·Y-PPase crystals in phosphate makes them crack [32]. In the Mn_2·Y-PPase·$(MnP_i)_2$ structure, the crystal contacts are different thus allowing the manganese phosphate to bind. Even so, the differences between the two yeast structures in this area are much smaller than the differences between the yeast and *E. coli* PPase structures (Fig. 3), leaving the E-PPase active site more open than that of Y-PPase. One effect of the structural changes occurring on product binding is to reduce solvent access to the active site somewhat. This is consistent with solution studies showing that binding of the third metal ion, firstly, changes the binding of P_i at P1 (see below) to protect Arg78 more effectively from chemical modification [30] and secondly, causes large changes in the ^{31}P chemical shifts of P_i in both P1 and P2 [33].

In the Mn_2·Y-PPase·$(MnP_i)_2$ structure, P1 forms hydrogen and ionic bonds to: Arg78 (bidentate coordination), Tyr192, Lys193, M3, M4 and three water molecules, one of which (Wat6) is coordinated to M3 (Fig. 5c). P2 forms hydrogen and ionic bonds to Lys56, Tyr93, M1, M2, M3, M4, and also with three water molecules: one (Wat1) bridging M1 and M2; one further hydrogen bonded to Tyr89; and one bridging M4 and Tyr192 (Fig. 5b). To our knowledge, P2 is the first reported P_i position bound to four divalent metal ions. The direct interactions of Lys56, Arg78 and Lys193 with P1 or P2 are consistent with solution studies showing the E-PPase variants mutated at the equivalent [24] positions all have an increased K_m [28] for Mg_2PP_i. The bond to Lys193 was not observed by Harutyunyan *et al.* [8].

Both Y-PPase [26,30] and E-PPase [16,31] bind the two P_is with substantially different affinities. Several lines of evidence indicate that P1 is the higher affinity site. Firstly, in the presence of three Mn^{2+} ions, P_i bound in the high affinity site protects Arg78 from chemical modification, and little additional protection is conferred by binding at P2 [30]. This can be explained by the bidentate coordination of the P1 phosphate with Arg78 (Fig. 5c). Secondly, the P1 site, but not the P2 site, is occupied by a sulphate ion in T-PPase [12]. Finally, one P_i ion binds to Y-PPase in the absence of divalent ion; its dissociation constant decreases a mere threefold when one or two

Figure 5



Product binding in Y-PPase. (**a**) An overall stereo view of the Mn_2·Y-PPase·$(MnP_i)_2$ active site showing most of the Lewis acid coordination of the product. Acidic residues are in red and basic in blue; tyrosine is in yellow; the two substrate phosphates are in orange; metals are in green; and waters in white. The view is rotated 180° from that of Figure 3 and residues Tyr93, Tyr89 and Lys154 have been removed for clarity. (**b**) The interactions around the P2 phosphate in a stereo view rotated 90° from the view in (a). Directly coordinated residues or atoms are in colour as described in (a), others are in white. Wat1 is coordinated to M1 and M2, and hydrogen bonded to Asp117 and to one of the phosphate oxygens (O1). (**c**) The interactions around the P1 phosphate, in a stereo view rotated 180° from (a). Colouring in the figure is as in (a). The proposed general acid, Wat6, is hydrogen bonded to O1 on P1 and is coordinated to M3. The bridging oxygen in the substrate pyrophosphate is proposed to be O3 on P1. (Figure generated using the programs MOLSCRIPT [70] and Raster3D [71].)

Mn^{2+} bind, but 30–80-fold when a third Mn^{2+} binds [30]. This pattern is fully consistent with binding to P1. At P1, two positively charged residues and a tyrosine hydroxyl offer a total of four sidechain hydrogen bond donors for binding (Arg78 is bidentate) and the P_i in P1 interacts with M3 and M4 but with neither M1 nor M2. By contrast, P2 has only one positively charged residue, Lys56, and a total of two sidechain hydrogen bond donors available for binding and P2 interacts directly with M1 and M2. P2 must also be the electrophilic P_i as it interacts directly

Table 1

Product–enzyme interactions in the Mn_2·Y-PPase·$(MnP_i)_2$ structure, within the substrate or between substrate and metal ion.*

					The P1 phosphate				The P2 phosphate			
	M1	M2	M3	M4	O1	O2	O3	O4	O1	O2	O3	O4
M1	-	3.70 (4.92)	5.63	3.72								2.46
M2		-	5.41	6.19					2.35			
M3			-	4.94	2.43					2.39		
M4				-				2.41				2.34
Wat1	2.40	2.56							2.71			
Wat6			2.47					2.97				

*All distances are in Å; distances in parentheses are those found for equivalent atoms in the Mn_2·Y-PPase structure. We used monomer 1 for both Mn_2·Y-PPase·$(MnP_i)_2$ and Mn_2·Y-PPase.

with the nucleophilic water and the electrophilic P_i leaves first [34] (see below).

The catalytic mechanism

The Mn_2·Y-PPase·$(MnP_i)_2$ structure (Fig. 5), when combined with the results of earlier functional studies in solution, permits for the first time a detailed proposal for the catalytic mechanism of PPase. PPase catalysis of PP_i hydrolysis is well described for both E-PPase and Y-PPase by the scheme shown in Figure 6. Indicating the precise conservation of mechanism, each of the microscopic rate constants, k_1–k_8, measured at pH 7.2 differs by less than threefold for the two enzymes [25]. There are four outstanding features of the scheme shown in Figure 6 (reviewed in [25]): the hydrolysis of PP_i proceeds by direct attack of water with no phosphorylated enzyme intermediate; the electrophilic P_i, having an oxygen from the nucleophilic water, is the first to dissociate from the enzyme; the equilibrium constant for hydrolysis of enzyme-bound PP_i, K_3 (k_3/k_4), is only five, in contrast to the hydrolysis of PP_i in solution which is strongly favourable; and PP_i hydrolysis and both P_i dissociation steps are all partially rate determining at pH 7.2.

Firstly, considering the hydrolysis step: k_3 at 25°C is 500–1000 s^{-1}, some 10^{10} faster than the corresponding rate of $MgPP_i$ hydrolysis in solution. Model studies of phosphoryl transfer in solution suggest three ways for catalyzing such transfer [35]. The first and most important is by lowering the pK_a of the leaving group, for which the Brønsted β is 1.2 [36]. The second is by shielding the charge on the electrophilic phosphorus, thus permitting attack by the stronger hydroxide anion rather than water. The third is by

Table 2

Product–enzyme interactions in the Mn_2·Y-PPase·$(MnP_i)_2$ structure.

Partner 1	Distance (Å)*	Partner 2	Distance (Å)*	Partner 3	Distance (Å)	Partner 4
K56 Nζ	2.88 (2.84)	E58 Oε2			2.89	P2 O3
E58 Oε2	2.48	M3				
K61 Nζ	2.87 (2.84)	D117 Oδ1				
D71 Oδ2	2.54 (3.13)	R78 Nη1				
R78 Nη1	2.73	P1 O1				
R78 Nη2	2.91	P1 O2				
Y93 Oη	2.96 (2.97)	D120 Oδ2	2.90 (2.91)	K154 Nζ	2.54	P2 O1
D115 Oδ1	2.44 (2.54)	M1				
D117 Oδ2	2.49	W1				
D120 Oδ1	2.31 (2.30)	M2				
D120 Oδ2	2.51 (2.47)	M1	2.68 (2.89)	K154 Nζ		
D147 Oδ1	2.81 (3.57)	K193 Nζ				
D147 Oδ2	3.36 (4.11)†	Y192 Oη	2.34	M4		
D148 Oε1	2.70	K193 Nζ				
D152 Oδ1	2.32 (2.39)	M1				
D152 Oδ2	3.09 (3.37)§	K154 Nζ	2.37	M4		
Y192 Oη	2.56	P1 O3				
K193 Nζ	2.75	P1 O2				

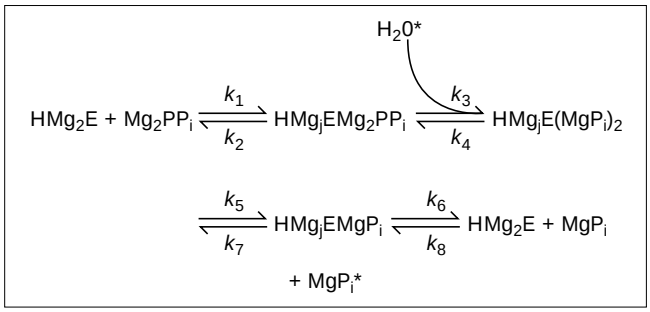
*Distances in parentheses are those found for equivalent atoms in the Mn_2·Y-PPase structure. We used monomer 1 for both Mn_2·Y-PPase·$(MnP_i)_2$ and Mn_2·Y-PPase. †There is no hydrogen bond between these two atoms in Mn_2·Y-PPase. §In Mn_2·Y-PPase the better hydrogen bond is from Lys154 Nζ to Asp152 Oδ1 (distance = 3.00 Å).

selectively stabilizing the trigonal bipyramid intermediate versus the tetrahedral reactant. Inspection of the Mn_2·Y-PPase·$(MnP_i)_2$ structure shows that Y-PPase is particularly well suited to catalyze phosphoryl transfer in at least the first two ways. Of the 22 lone pairs on the two product phosphates, 18 are coordinated: 6 by hydrogen bonds to protein sidechains, 6 by hydrogen bonds to water, and 6 by coordination to metal ions.

We were able to generate two types of models for the catalytic mechanism of the third step. The first model required two water molecules: Wat1 (Fig. 5b), the putative bridging hydroxide served as a general base to activate a second water as the incoming nucleophile. The second water molecule was placed between O1 and O2 in the product complex (Fig. 5b) (see below). In the second model, which we favour because there was less steric conflict in the active site, Wat1 is itself the attacking nucleophile. This model is shown schematically in Figure 7. It was generated by docking a model of the transition state into the structure of the product complex (Fig. 5a) without making any changes to the protein and trying to overlap the phosphorus oxygens in the transition state and product as much as possible. Total overlap was not possible, because an inversion of configuration is in progress at P2, and because no oxygen on P1 in the product complex points directly towards P2. Nonetheless, the only adjusted parameters, all in the transition state, were the axial O–P bond lengths on the trigonal bipyramidal P2 (range 1.9–2.4 Å), the P–O–P bend angle (range 130–180°), and the torsion angles in the transition state. In the model, 15 of the 17 lone pairs on the substrate PP_i are coordinated, five by metal ions and ten by hydrogen bonds (Fig. 7).

In this model, we assign the phosphoryl groups in P1 and P2 as the leaving and electrophilic groups, respectively. The basis for this assignment is that the electrophilic P_i dissociates before the leaving group P_i [34], that the more

Figure 6



The minimal scheme of PPase catalysis, where j = 1 or 2, depending on pH and divalent metal ion concentration. (In general E-PPase requires one more metal ion for maximal activity at low pH than Y-PPase.) (Details are given in the text).

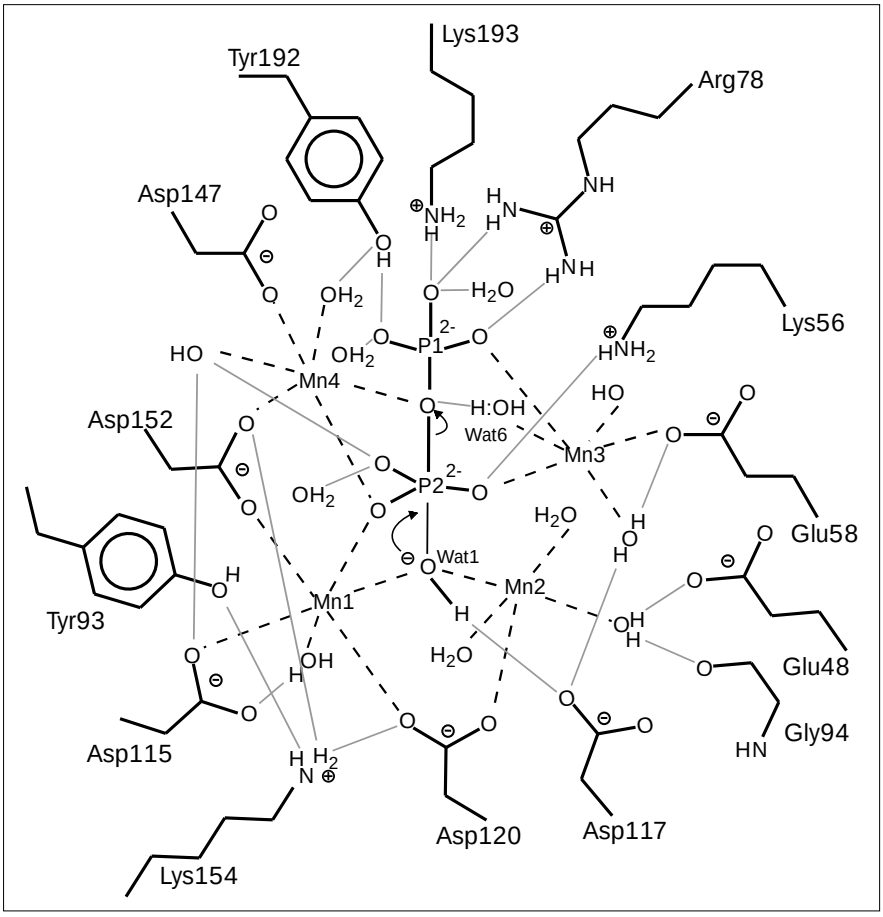
weakly bound P_i should dissociate first, and that P1 is the high affinity site. Furthermore, P2 is closer to the only reasonable candidate for the general base/nucleophile, Wat1 (Fig. 5a). Placing the leaving group in the high affinity site also makes sense mechanistically, as tighter binding should correlate with higher conjugate acidity and thus a faster hydrolysis rate. The pK_a of P_i in P1 should be lower than that of P_i in P2, as strong hydrogen bonding should lower pK_a more effectively than coordination to metal ions, especially coordination to relatively weak acidic metal ions like Mg^{2+} and Mn^{2+} [35,37,38].

Although our assignment of P2 as the electrophilic site agrees with Harutyunyan *et al.* [8], we disagree with their assignment of P2 (their P_iA) as the high affinity site and with P1 (their P_iL) as the first phosphate to leave the enzyme. Their assignment is completely inconsistent with the solution data referred to above [34], and apparently was made because P2 looks like it is more buried than P1 (Fig. 5a). However, a space-filling model (Fig. 8) shows clearly that both are exposed to solvent. P2 is also well positioned to leave after the chemical catalysis step as it can dissociate with the exposed M3 metal ion; such dissociation may be facilitated by the deep groove that leads down to the P2-binding pocket in the active site (Fig. 8).

The pH dependence of k_{cat} for PPases indicates the presence of an essential general acid, of apparent pK_a of 8.5 (Y-PPase) [39] to 9.7 (E-PPase) [31]. We consider it likely that the general acid protonates P1 during the hydrolysis reaction, thus further lowering its pK_a. Arg78, Lys193 and Tyr192, although all potentially general acids because each donates hydrogen bonds to P1, are unlikely candidates because variants of E-PPase, in which the corresponding residues [24] have been mutated, retain appreciable activity and show little change in the apparent pK_a of the essential acid. We therefore believe the general acid is a water molecule coordinated to P1. Three water molecules (Wat6, Wat119 and Wat182) (Fig. 5c) are directly coordinated to P1 in the Mn_2·Y-PPase·$(MnP_i)_2$ structure. The pK_as of Wat119 and Wat182 are unlikely to be changed much from the normal value of 16, as they are exposed to solvent at the top of the active site and are coordinated neither to protein residues nor to metal ions. Wat6, on the other hand, is buried to a greater extent and, because it is coordinated to M3, is more likely to have a lowered pK_a. In addition, Wat6 is well positioned to donate a proton to an oxygen in P1.

The mechanistic model produced by docking (Fig. 7) implies that the Mn_2·Y-PPase·$(MnP_i)_2$ structure (Fig. 5a) has relaxed since the chemical catalytic step. The oxygen from the putative nucleophilic hydroxide anion (Wat1 in Fig. 7; bridging M1 and M2) has rotated away to become coordinated to M2 alone in Figure 5b and another water molecule has therefore replaced it in bridging M1 and M2

Figure 7



The proposed PPase mechanism. The active-site coordination is shown in two dimensions; hydrogen bonding is shown with grey lines and metal coordination with dashed lines. The flow of electrons is indicated with arrows. The reaction involves a hydroxide ion (Wat1), coordinated to metals Mn1 and Mn2, which hydrolyzes the substrate P_2O_7–Mn_2. P1 and P2 are labels for the leaving and electrophilic phosphoryl groups, respectively. The required general acid is a water molecule (Wat6) coordinated to Mn3 and hydrogen bonded to P1 in the Mn_2·Y-PPase·$(MnP_i)_2$ structure. (The two-dimensional presentation of the mechanism was initially created using LIGPLOT [72] and edited using CorelDRAW 4.0.)

(Fig. 5b). The Mn_2·Y-PPase·$(MnP_i)_2$ product complex structure we observe is thus a hydrated form of the product of the hydrolysis step of the reaction (Fig. 6).

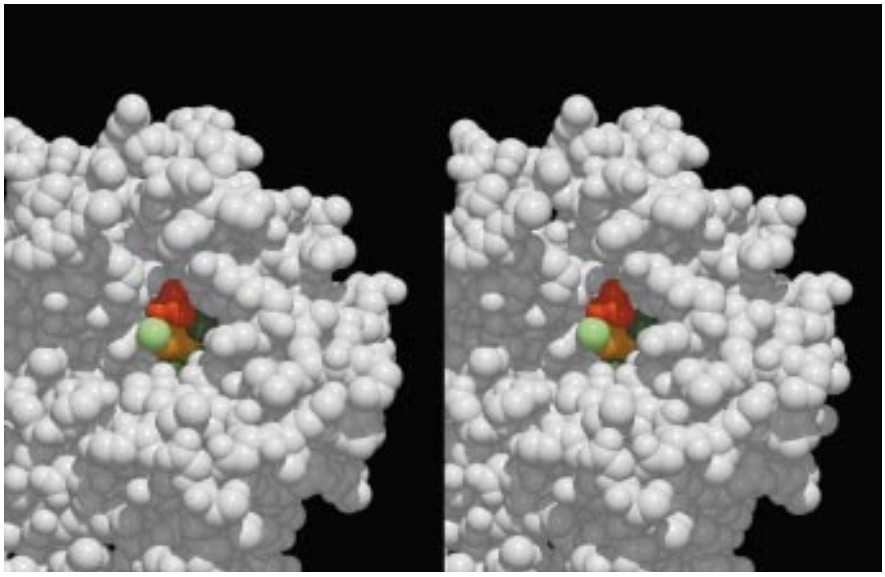
This proposed nucleophilic role for Wat1 in Figure 7 is in accordance with recent rate studies of ours. The pH dependence of k_3 for mutant forms of E-PPase [16,31] (Fabrichniy *et al.*, unpublished data) provides evidence for an essential base, with an apparent pK_a between 7.8 and 8.7 for the active-site variants Asp97→Glu, Tyr55→Phe, and Lys104→Arg (corresponding to Asp147, Tyr93, and Lys154 in Y-PPase). The corresponding pK_a in wild-type E-PPase is <6.5. Measured by the effect on k_{cat}, conservative mutation of virtually every conserved active-site residue in E-PPase increases the apparent pK_a of the essential base by 1–3 pH units [28]. While this latter result is more difficult to interpret precisely, because the pH dependence of k_{cat} reflects the pH dependence of k_3, k_5, and k_7, clearly mutation at residues throughout the active site increases the pK_a of an essential base. We reasoned that the essential base/nucleophile had an abnormally low pK_a that depended on the integrity of the active-site structure, so that any perturbation of that structure increased its pK_a. We therefore proposed [28] that the essential base was an hydroxide, which is stabilized by binding to one or two metal ions at the active site, and perhaps by binding to a lysine or arginine residue as well. Consistent with this proposal, Wat1 is bound to M1 and M2, thus reducing its pK_a. A further reduction might be possible through hydrogen bonding to Asp117.

Clearly Wat1, which may even partially resemble an O^{2-} anion if proton transfer to Asp117 occurs, is a very potent nucleophile. However, because of its negative charge it can only attack P2 (the electrophilic phosphorus) if the negatively charged oxygens on P2 are charge-shielded [4,40]. This requirement is met for the phosphoryl group in P2, which binds to all four divalent metal ions and to Lys56. Metal ion binding also appears to stabilize the pentacoordinate trigonal bipyramid intermediate in other phosphoryl transfer enzymes [41,42].

Two aspects of the PPase catalytic mechanism remain unclear. The first is the timing of proton transfer to the leaving group phosphate. We have proposed in Figure 7 that the transfer is synchronous with P–O–P bond breaking,

Figure 8

A space-filling stereo view of the active site of Y-PPase in the Mn_2·Y-PPase·$(MnP_i)_2$ structure, showing how exposed the two product phosphates are. Bound water molecules have been removed for clarity. P1 is shown in red, P2 in orange, M3 in light green and the other metals in dark green; the protein is shown in white. Arg78 binds in a bidentate fashion to P1; its headgroup is visible to the left and above P1. The channel leading down to P2 can be seen to the left of M3, explaining how the M3–P2 complex can dissociate first. (Figure generated using the programs MOLSCRIPT and Raster3D.)



but this is by no means certain. Protonation of the leaving group in the transition state, to lower the energy barrier and increase k_3, may not be necessary; the combination of hydrogen bond donors and metal ions binding to P1 may make it a good enough leaving group (i.e. have a low enough pK_a). In this case, protonation would take place after hydrolysis, as the dominant species in solution at neutral pH is $MHPO_4$ (not MPO_4^- as indicated by Harutyunyan *et al.* [8]).

The second unclear aspect of the mechanism is the identity of the nucleophile. The current mechanistic model has the bridging hydroxide corresponding to Wat1 as the nucleophile, as this model can be built most easily into the active site. This is similar to the models proposed for other enzymes that have a similar metal–oxygen–metal arrangement (see below). However, it does imply that there is a conformational change after chemical catalysis, as the water (Wat1 in the product complex Mn_2·Y-PPase·$(MnP_i)_2$ structure [Fig. 5]) is obviously not one of the product P2 oxygens. (The O1 oxygen on P2 [Fig. 5b] is assumed to have come from another water molecule in the Wat1-binding site in the transition state [Fig. 7].) It is also possible to build a model based on the assumption that the Mn_2·Y-PPase·$(MnP_i)_2$ structure accurately reflects the product of the hydrolysis step (Fig. 6) as appears to have been done by Harutyunyan *et al.* [8]. In this latter model the nucleophilic water is placed most plausibly between the positions of O1 and O2 depicted in Figure 5b, and thus its oxygen corresponds to either O1 or O2. This nucleophilic water could form a hydrogen bond with Wat 1; in the direction of PP_i hydrolysis, it is activated as a nucleophile by proton removal via hydrogen bonding to the hydroxide corresponding to Wat1. Coordination to either M2 or M3 would then aid this process. The major disadvantage of such a model, versus the model presented in Figure 7, is that it is much harder to satisfy the geometrical requirements for catalysis without either distorting the transition state beyond acceptable limits or invoking conformational change on the part of the protein. Further solution and structural studies will be required to address these questions, as well as to examine critically the specific roles in catalysis assigned to amino acid residues at the active site.

Catalytic roles of the fourth divalent metal ion

At neutral pH E-PPase requires four Mg^{2+} ions per active site for activity, whereas the yeast enzyme is active with either three or four Mg^{2+}, with the three Mg^{2+} form having somewhat higher activity. Our structural studies provide an interesting rationale for this difference, based on two opposing effects on k_{cat} attributable to the binding of the fourth divalent metal ion.

The first effect at neutral pH on k_{cat} is the deprotonation of the essential basic group modulating k_{cat}, (i.e. Wat1 bridging M1 and M2). In E-PPase, the pK_a of this group rises from 6.7 with four Mg^{2+} ions bound to 8.7 with three Mg^{2+} ions bound [16]; at neutral pH the three Mg^{2+} form of the enzyme is thus quite inactive. Although a full analysis of k_{cat} as a function of pH and Mg^{2+} concentration has yet to be carried out for Y-PPase, the pK_a of Wat1 with three Mg^{2+} bound to Y-PPase appears to be low enough so that it is substantially deprotonated at neutral pH [39,43] (Pohjanjoki *et al.*, unpublished data). Binding of a fourth Mg^{2+} ion is thus not required for deprotonation of Wat1 in Y-PPase. This suggests that the interactions that lower the pK_a of Wat1 are stronger in Y-PPase than in E-PPase. Consistent

with this, the Y-PPase active-site cleft is slightly more closed than the E-PPase cleft, and the distance in the Mg·E-PPase structure from M1 to the putative position of M2 is 6Å (data not shown), versus the measured value of 4.9Å in the Mn_2·Y-PPase structure. Although this might be due to the different metallation state of the enzyme in the crystals, it is more likely due to the dif-ference in the β4–β5 loop region of the two enzymes (Y-PPase, 101–115; E-PPase, 63–65). The shorter E-PPase connection probably prevents M1 from moving as much towards M2 in E-PPase as it does in Y-PPase.

The second effect at neutral pH is on the magnitude of k_7, which decreases on binding of the fourth Mg^{2+} ion to Y-PPase. It is this decrease, rather than any effect on k_3, that accounts for the small inhibitory effect of excess Mg^{2+} on Y-PPase activity [26]. A structural rationale for this effect is provided by the observation that the mobile loop (144–152) is drawn in more closely in the Mn_2·Y-PPase·$(MnP_i)_2$ structure than in the Mn_2·Y-PPase structure (Fig. 2). This change must be due, at least in part, to the binding of M4 to Asp147 and Asp152, and should slow P_i release from P1, thus decreasing k_7. Both effects are operative in E-PPase, with the effect on the pK_a of Wat1 being more important, and so binding to M4 increases k_{cat}. In Y-PPase, only the second effect is important and so binding to M4 decreases k_{cat}.

The structure further suggests explanations for some of our other kinetic data obtained with E- and Y-PPase. Firstly, the large increase in k_7 for E-PPase as the pH is increased from 6.5 to 9.3 [16] may be due to deprotonation of Tyr192, to which M4 is coordinated through a water molecule (Fig. 5a). This would accelerate P1 dissociation by charge repulsion, overwhelming the effect on k_7 mentioned above. Secondly, not only can a fifth divalent metal ion bind to E-PPase [16] but also such binding is required for efficient catalysis in some E-PPase active-site variants [31,44]. This may be because the E-PPase active site is more open (Fig. 3; see above). Conversely, the more closed and presumably more rigid Y-PPase active site also explains why active-site variants in Y-PPase are generally less active than the equivalent variants in E-PPase [27].

Pyrophosphatase and its relationship to other phosphoryl transfer enzymes

PPase, ATPase, and kinases catalyze the following reaction:

$$(R)OPO_2^-OPO_3^{2-} + R'–OH \rightarrow R'OPO_3^{2-} + (R)OPO_3^{2-} + H^+$$

(For PPase, R = H^+ or a negative charge and R′ = H; for ATPase, R = adenosyl 5′–OPO_2^- and R′ = H; and for kinases, R = adenosyl 5′–OPO_2^- and R′ = various.)

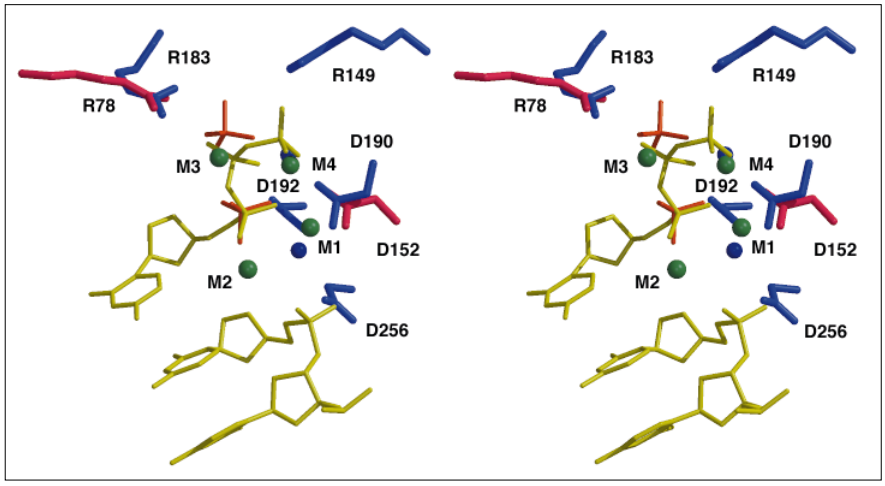
These enzymes thus solve very similar chemical problems and so could have very similar catalytic mechanisms. In fact, this is not the case, Nature having evolved at least two different ways of catalyzing this reaction. As demonstrated above, PPase solves the problem by activating the attacking water as a nucleophile and by the massive use of Lewis acids. In particular, three to four divalent metal ions are used both to stabilize the leaving group P_i and to charge-shield the anionic electrophilic phosphoryl group from attack by the anionic nucleophile. ATPases and kinases, in contrast, employ a different strategy. They generally have a single divalent cation involved in catalysis and make extensive use of amide backbone protons, particularly those from a set of highly conserved glycines, as Lewis acids [45]. As a result, there is no obvious structural or evolutionary linkage between PPases on the one hand and ATPases and kinases on the other. Such a linkage was proposed by Baltscheffsky and coworkers [46], based on a short conserved sequence motif, found in PPases and the α and β ATPases from *Rhodospirillium rubrum*, *Rhodospirillium blastica* and *E. coli* and the fact that PPases, like ATPases, can catalyze the synthesis of a pyrophosphate linkage. This sequence (114-Gly-Asp-Asn-Asp-Pro-Ile-Asp-120 in Y-PPase) includes three of the important PPase aspartates. Using the bovine F_1-ATPase coordinates [47] we found no evidence of structural similarity between the two proteins in this region (110-Ala-Leu-Gly-Asn-Ala-Ile-Asp-116 in F_1-ATPase); nor is this region part of the active site of F_1-ATPase. Clearly, this throws some doubt on the use of very short sequence motifs like those found in ProSite [48] as diagnostic tools in the absence of any other information.

To what other proteins then, if any, is PPase similar? Searches run by G Kleywegt using his program SPASM found many protein metal binding sites with similar local structure, but the similarity never extended beyond the metal ion binding carboxylate groups themselves. However, there is an interesting and somewhat unexpected similarity to proteins which interact with DNA. PPases form a superfamily (i.e. presumed to be unrelated by divergent evolution to anything else) in the OB-fold defined in the SCOP database [49]. The distinguishing feature of this fold is a five-stranded β barrel with a shear number of eight. Two other OB-fold proteins (gene V protein of bacteriophage f1 [PDB code 1BGH] and cold shock DNA-binding protein [PDB code 1MJC]) bind DNA on the top surface of their β barrels, reminiscent of the location of the active site in PPase.

A perhaps more interesting similarity is with the active sites of some polymerases. Pelletier *et al.* [50] have proposed a structure-based mechanism for rat DNA polymerase β (pol β). The mechanism is, it should be noted, different from that of the Klenow fragment of *E. coli* DNA polymerase I [18], HIV reverse transcriptase [20], and a thermostable *Bacillus* spp DNA polymerase (L Beese, personal communication). The structure of pol β [51] contains two Mg^{2+} ions which align with the PPase active site M1 and M4 (Fig. 9). The metal ions are both coordinated to one aspartate, Asp152 in the yeast PPase and Asp190 in pol β. Asp→Glu

Figure 9

A stereo view of the superposition of rat DNA polymerase β (polβ) (in purple) and Y-PPase (in red). The magnesium ions (pol β) are in purple; the manganese ions (Y-PPase) are in green; the pol β substrate is in yellow and the Y-PPase product is in orange. The lower metals in both enzymes take part in activation of the nucleophile and the upper metal coordinates the leaving group. Arg78 in Y-PPase and Arg183 in rat polβ are involved in substrate activation. (Figure generated using the programs MOLSCRIPT [70] and Raster3D [71].)

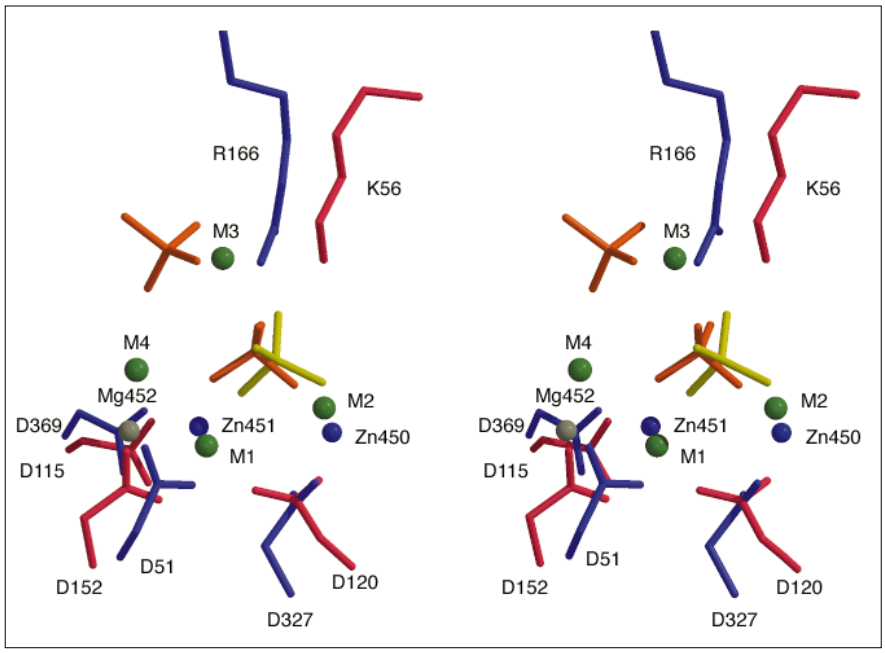


mutations of these residues have similar effects on activity in both yeast PPase and pol β [32,52]. In pol β, the 'lower' metal ion in Figure 9 is believed to be involved in activating the attacking nucleophile, the 3′-OH of the primer strand; in PPase, M1 is similarly involved with Wat1. The leaving group activation also seems to be similar; both substrates are coordinated by positive residues and by a metal ion (Fig. 9). In pol β, the leaving group is larger (PP_i instead of P_i) and there are two arginines, instead of just one, but the orientation of the head group of Arg183 is very similar to that of Arg78 in the Mn_2·Y-PPase·$(MnP_i)_2$ structure (Fig. 9). The two other conserved aspartic residues in pol β (Asp192 and Asp256) do not have direct counterparts in PPases but Asp256 helps remove the proton from the 3′-OH as does Asp117 in Y-PPase, but from a different orientation [51]. The direction of attack in the two enzymes is also different. In Figure 9, the Y-PPase Wat1 (not shown) is in front of M1 and M2, whereas the primer 3′-OH in pol β would be behind the equivalent of M1.

In the alignment of PPase with the 3′,5′-exonuclease active site of Klenow fragment (Klenex) (data available with

Figure 10

A stereo view superposition of alkaline phosphatase (in purple) and Mn_2·Y-PPase·$(MnP_i)_2$ (in red). The Y-PPase product is in orange and the alkaline phosphatase P_i is in yellow; Mn^{2+}, Zn^{2+} and Mg^{2+} ions are in green, purple and grey, respectively. Despite the locally similar arrangement of active-site atoms, the mechanism is likely to be different (see text).



supplementary materials), only the active site including the putative hydroxide ion aligns well, as with pol β. The arrangement of the metal ions (M1 and M4 in Y-PPase) with the putative hydroxide ion is essentially identical in the two enzymes. The geometry around the tightly bound metal ion (PPase M1; Klenex Zn930) is different, however, as the zinc ion has trigonal bipyramidal geometry, thus rationalizing the preference of this site for zinc over magnesium [18]. The preference for the more acidic Zn^{2+} ion explains why Klenex only requires one metal ion (Zn930) to generate the hydroxide ion, whereas PPase requires two (M1 and M2). Both binding sites include three carboxylate groups: two aspartates and a glutamate, in Klenex, three aspartates in PPase. The coordination around the other metal ion (PPase M4; Klenex Zn931) is in both cases octahedral and the site preferentially binds Mg^{2+} [18]. Klenex Zn931 makes a bidentate interaction with the 5′-phosphate of the thymidine 5′-phosphate (TMP) in the active site. The phosphate on the TMP corresponds to, and aligns quite well with, our P2; it is the electrophile. Viewed from the perspective of the electrophile, in both enzymes a single oxygen on the phosphate is coordinated by two metal ions. The putative position of the leaving group nucleotide 3′-hydroxyl, not present in the Klenex structure, would be approximately the same as P1. As the bidentate coordination of Zn931 to the 5′-phosphate of the TMP appears quite strained, its coordination may be monodentate to the phosphate and monodentate to the hydroxyl when the leaving group is bound. If so, the coordination around Zn931 would mirror that of M4 even more closely.

PPase also has an active site similarity with alkaline phosphatase (AP). Kim and Wyckoff [21] solved the structure of alkaline phosphatase (AP) with one Mg^{2+} and two Zn^{2+} ions in the active site and proposed a two metal ion mechanism using the Zn^{2+} ions. The Zn^{2+} ions (451 and 450) and the phosphate in AP can be aligned to the Y-PPase M1 and M2 and phosphate P2. The carboxylate groups of Asp115, Asp120 and Asp152 in Y-PPase superimpose on the AP Asp369, Asp51 and Asp327 carboxylates and Lys56 superimposes with AP Arg166 (Fig. 10). Despite the similar arrangement of residues and metals in the active site, the overall structures of PPase and AP show no similarity and seem to have evolved separately, as is also true for AP and polymerases. Furthermore, the mechanisms of alkaline phosphatase and PPases differ in two significant ways. AP has a covalent enzyme-bound phosphoserine intermediate [21] and requires the more acidic Zn^{2+} ion, whereas PPases catalyze the reaction by direct hydrolysis [14] and the rate constant k_3 (Fig. 6) is approximately the same [26,53] whether Mg^{2+} or Zn^{2+} is the cofactor.

Are PPases related to DNA polymerases and AP by more than just the mechanistic similarity already noted by others [22]? Nature often appears to find multiple structural solutions that correspond to the same catalytic machinery. For catalytic triad hydrolysis, five families are known, all apparently related by convergent evolution: serine proteases, the αβ-hydrolase fold [54], cysteine proteases, subtilisins, and the recently identified N-terminal nucleophile (NTN) hydrolases [55]. Similarly, for phosphoryl transfer, at least four different solutions exist, reflecting how ancient and critical a reaction it is. These include ATPases and kinases, which use the giant anion hole, and three types of catalysis which require multiple divalent metal ions: the three-metal mechanism of soluble inorganic pyrophosphatases, the two-metal mechanism of the nucleic acid polymerases and exonucleases, and the covalent intermediate mechanism of AP. However, as these enzymes are unrelated by divergent evolution, we believe the similarities in the mechanism described above merely reflect the limited range of catalytic options available to enzymes catalyzing phosphoryl transfer.

Implicit in the statement that phosphoryl transfer has evolved more than once is the idea that enzyme-catalyzed phosphoanhydride hydrolysis must be a very old reaction. Indeed, it must have been present as organisms moved from the 'RNA world' to the 'DNA world', and possibly even as proto-organisms moved from the 'PP_i world' to the 'ATP world' [56].

Biological implications

Inorganic pyrophosphate (PP_i) is produced by many biosynthetic reactions, such as protein, RNA, DNA and polysaccharide biosynthesis. These reactions are driven to completion by the essentially irreversible hydrolysis of PP_i, catalyzed by soluble inorganic pyrophosphatase (PPase) [3]. Not surprisingly, therefore, PPases are ubiquitous and essential, as the very high levels of PP_i that would result from the absence of PPase would certainly be toxic. For example, Klemme [57] calculated that the internal concentration of PP_i would rise to 3M in one hour in rapidly growing *Escherichia coli* cells, in the absence of PPase (the actual concentration is about 1mM).

The mechanisms of phosphoryl transfer enzymes remain incompletely understood [4]. In addition to its importance in the control of overall phosphorus metabolism, PPase provides a useful model of a phosphoryl transfer enzyme with a complex but well understood dependence on metal ion concentration and pH [16]. Thus progress towards a complete catalytic model of PPase, combining structural, enzymological and mutagenesis studies, holds out the prospect of increasing our understanding of two important issues: the way that metal ions are used to effect hydrolysis in biological systems, and the various ways that biological systems catalyze phosphoryl transfer.

We present two structures of the PPase from the budding yeast *Saccharomyces cerevisiae* (Y-PPase), one of the activated enzyme (the Mn_2·Y-PPase complex) and the

other of the product complex (Mn_2·Y-PPase·(MnP_i)$_2$). These structures, combined with the results of functional studies of the wild type and variant enzymes, permit us to formulate a substantially complete structure-based model for the PPase mechanism, and to compare and contrast this model with mechanisms proposed for other phosphoryl transfer enzymes.

The mechanistic model has three novel features. Firstly, there is a highly charged active site in which virtually every lone pair in the product, substrate and transition-state is coordinated by metal ions, by hydrogen bonds to protein sidechains, or by hydrogen bonds to ordered water molecules. This achieves both charge shielding of the electrophilic phosphoryl group (P2 in Fig. 5b) and activation of the leaving group (P1). Secondly, the model identifies an hydroxide ion nucleophile (Wat1), which is stabilized because it bridges two metal ions (M1 and M2; Fig. 5b) and is hydrogen bonded to Asp117; the attacking species may thus have partial O^{2-} character. Thirdly, it identifies a water molecule stabilized by coordination to M3 as a likely candidate for the general acid protonating the leaving group phosphate (P1; Fig. 5c). Thus at least three metal ions are required for catalysis.

Though the mechanism and structure is conserved among PPases, no other phosphoryl transfer enzymes seem to be related to them by divergent evolution. We thus call this mechanism the 'three-metal ion mechanism', in contrast to the 'two-metal ion mechanism' proposed for alkaline phosphatase, polymerases and exonucleases, among others. If PPases are not related by divergent evolution to these other phosphoryl transfer enzymes, the common catalytic motifs employed in enzymatic phosphoryl transfer must have evolved more than once. We believe this reflects how old enzyme-catalyzed phosphoanhydride hydrolysis is: it must have been present as organisms moved from the 'RNA world' to the 'DNA world'.

Materials and methods

Crystallization and data collection

Protein purification was as described in Heikinheimo *et al.* [27]. We chose Mn^{2+} for crystallizations instead of the physiological Mg^{2+}, because it binds more tightly [58] and because Mn^{2+} can be easily located by anomalous difference Fourier (Fig. 1).

The Mn_2·Y-PPase crystals were grown as described elsewhere [32]. The crystals were soaked overnight in 9% PEG 1000, 11% PEG 4000, 50 mM MES, pH 6.3, 1 mM $MnCl_2$, mounted in the same solution, and data were collected at 0° C. The Mn_2·Y-PPase crystals belong to the space group $P2_12_12_1$ and have unit cell dimensions of 54.3×68.4×161.9 Å, with two monomers in the asymmetric unit. The Mn_2·Y-PPase·(MnP_i)$_2$ crystals were grown in 8 µl sitting drops. The conditions for growth were 30 mM MES, pH 6.0, 1 mM $MnCl_2$, 1 mM sodium hydrogen phosphate with a drop 2-methyl-2,4-pentanediol (MPD) concentration of 17% and a well concentration of 19%. Crystals appeared in 3–5 weeks from the set-up of the drops and typically grew to a size of 0.3×0.4×0.6 mm. Crystals were mounted with 19% MPD, 30 mM MES, pH 6.0, 1 mM $MnCl_2$, 0.5 mM sodium hydrogen phosphate and soaked in this solution for a few hours prior to mounting. Mn_2·Y-PPase·(MnP_i)$_2$ belong to the space group $P2_12_12_1$ and have unit cell dimensions of 59.1×103.9×117.3 Å with two monomers in the asymmetric unit. Data from these crystals were collected at –15° C. Data from both crystal forms were processed with the program Denzo [59] (Table 3).

Molecular replacement and structure refinement

Molecular replacement and refinement were done in X-PLOR [60] using modified dictionaries of protein structure parameters [61]. We used the program O [29] for model building. ($3F_o–2F_c$) and ($F_o–F_c$) maps calculated by X-PLOR were used for model building and simulated-annealing omit ($F_o–F_c$) maps were used several times in the most difficult parts of the molecule. The metal-binding sites were identified from anomalous difference maps calculated with PROTEIN [62]. The Mn_2·Y-PPase structure gave two anomalous peaks per monomer whereas the Mn_2·Y-PPase·(MnP_i)$_2$ structure gave four anomalous difference peaks per monomer.

Waters were added to the protein models, from a resolution of 2.5 Å onwards, to 4σ ($F_o–F_c$) difference map peaks, picked with MAPMAN (in RAVE) [63] and refined to the difference map with an O macro created with MOLEMAN [64]. After this, all peaks were checked by hand, and only those with acceptable hydrogen bonding were accepted. At 2.5 Å we also started grouped B factor refinement.

The Mn_2·Y-PPase structure

The structure was solved by molecular replacement using data from 15 to 4 Å, using the partially refined apoY-PPase coordinates kindly provided by Professor Voet (University of Pennsylvania). After the rotation and translation searches, the starting R factor was 45%. 10% of the reflections were set aside for R_{free} calculations [65]. All the electron-density maps were, however, calculated using the full data set. We extended the resolution to 3 Å by adding 10% more reflections at a time and then doing rigid-body refinement after each addition. At 3 Å, the free and working R factors were 42.1 and 35.6%, respectively. This model was then refined with the slowcool routine in X-PLOR [60], after which the R factors were 37.3 and 26.0%. Omit maps of the Mn_2·Y-PPase model showed three regions that were somewhat incorrect. The loops 34–43 (β-1 to β1) and 231–240 (following αB) at the surface of the model were very different. Residues 60–71 in the interior of the protein (including sheet β3) were incorrect as they are in the 1PYP coordinates in the PDB. This might have been due to an error in the original peptide-derived protein sequence, which omits Lys74 [23]. Having this loop one residue short, the partially refined apoY-PPase structure returns to the correct tracing.

The two monomers were restrained to be similar during the refinement with a non-crystallographic symmetry (NCS) weight of 300. At 2.5 Å resolution (R factors 32.3 and 25.8%) the crystal contact areas were released to be different, but the core was kept similar. At the same time, grouped B factor refinement was started, after which the R factors were 29.8 and 22.7%. After further model building and refinement, the R factors converged at 28.5 and 21.2%, so waters were added as described above. The B factors were refined as grouped Bs [60], where only two B factors are allowed per residue. Finally, 25 cycles of individual B factor refinement were run with X-PLOR bsigma restraint parameters of 1.5 and 2.5 for mainchain and sidechains, respectively, and asigma restraint parameters of 2.0 and 2.5. This lowered the R factors from 24.9 and 18.5% to 24.3 and 17.2%.

The Mn_2·Y-PPase·(MnP_i)$_2$ structure

The Mn_2·Y-PPase·(MnP_i)$_2$ structure was solved by molecular replacement from the Mn_2·Y-PPase structure. Refinement was essentially as described above. During the slowcool runs using data from 8.0 to 3.0 Å the two monomers in the asymmetric unit were constrained to be identical. As we increased the resolution, it became apparent that the loop 72–76 was packed differently in the two monomers and had to be released. The rest of the molecule was restrained to be similar with a

Table 3

Data collection and refinement statistics.

	Mn_2·Y-PPase	Mn_2·Y-PPase·$(MnP_i)_2$
Number of crystals	2	8
R_{merge} (I)* (%)	8.8	10.3
Resolution (Å)	8.0–2.2	8.0–2.0
Independent reflections (total observations)	27 467 (178 216)	45 565 (236 904)
Completeness (%) (last resolution shell)	89.5 (2.2–2.3 : 71.9)	89.2 (2.0–2.09 : 82.9)
Reflections $\|F\| > 2\sigma$ (F)	24 743	40 088
R_{free} setup (%)	10	4.3
Number of reflections	2 777	2 263
R_{work}/R_{free}§ (%)	17.2/24.3	17.2/20.5
Waters	195	307
Number of non-hydrogen atoms	4589	4789
Average temperature factors (Å^2)		
protein	27.8	17.7
phosphate ions		9.7
manganese ions	22.6	13.5
water molecules	31.4	25.0
Ramachandran		
most favoured (%)	89.5	89.0
Rms deviation from ideality		
bond lengths (Å)	0.010	0.010
bond angles (°)	1.6	1.6
dihedral angles (°)	25	25

*$R_{merge} = \Sigma|I_h - <I>| / \Sigma I_h$, where I_h is the measured intensity and $<I>$ the average intensity of reflection hkl. §R_{free} and R_{work} are conventional crystallographic R factors; $R = \Sigma||F_{obs}| - |F_{calc}|| / \Sigma|F_{obs}|$, where F_{obs} is the observed and F_{calc} the calculated structure factor of reflection hkl. R_{free} is calculated on the 'test set' of reflections, as described in the X-PLOR manual [55].

very high weight (500) until 2.2 Å (R factors of 23.1 and 20.2 %), when the weight was lowered to 300 and some of the sidechains having different conformations were allowed to be different. At 2.2 Å (R factors 22.8 and 18.6 %) the NCS restraints were removed and data added to 2.0 Å in four steps of positional refinement. At 2.0 Å more waters were added and individual B factors refined. The final R factor for this structure was 17.2 % and the R_{free} was 20.5 %.

For building the more restrained Mn_2·Y-PPase·$(MnP_i)_2$ structure we averaged the maps using the program AVE (in RAVE [63]). We read the rotation matrix out from O and created the mask for the protein in MAMA [66]. We were able to trace the first 282 residues in the Mn_2·Y-PPase·$(MnP_i)_2$ structure without chain breaks. In the Mn_2·Y-PPase structure the C terminus in monomer 1 is partly packed between the molecules in the crystal and we have traced it to residue 283. The loop 234–238 of the second monomer in the Mn_2·Y-PPase structure is disordered and thus omitted from the final model. An example of the final refined models and there fits to the electron density is shown in Figure 1.

The differences between the two monomers in the asymmetric unit are smaller in the Mn_2·Y-PPase·$(MnP_i)_2$ structure than in the Mn_2·Y-PPase structure due to the tight NCS restraints imposed during the refinement (see above). All Cα atoms in these two monomers superimpose with an overall rms deviation of 0.18 Å. For the two monomers in the Mn_2·Y-PPase asymmetric unit, 268 residues superimpose with an overall rms deviation of 0.20 Å. This provides an alternate measure of the accuracy of the structures, which is around 0.2 Å.

Structure analysis

The structures were analyzed with the programs PROCHECK [67] and WHAT CHECK [68] and the corrections proposed by the programs were taken into account during the last rounds of refinement. Structure superpositions were done using three programs: RgbSup (P Fitzgerald, unpublished program), SPASM (G Kleywegt, unpublished program) and O [29]. All PPases were superimposed with RgbSup, in which the parts of the proteins that differ by more than a specified cut-off are removed in a cyclic fashion from the superposition. It is thus more likely to detect small differences between protein structures. We used a cut-off of 0.5 Å for superimposing Y-PPases on each other and 1.0 Å to align Y-PPases to E-PPases. In the Y- and E-PPase superpositions, the superposition was first done in O and then refined with RgbSup.

We also sought to identify proteins related to PPase by searching the Hohobom and Sander protein structure database (version 1995) [69] for a three-dimensional arrangement of carboxylate residues similar to the five aspartates (Asp115, 117, 120, 147 and 152) important for metal-binding and function. We used the program SPASM (G Kleywegt, unpublished program) which looks for a similar three-dimensional arrangement of carboxylates, without constraints on the amino acid identity or sequence order. Both the full set of these five aspartates as well as smaller subsets were employed in the search.

Alkaline phosphatase (AP) (1alk) [21], the rat DNA polymerase β (pol β) (2BPF) [50] and the Klenow fragment (Klenex) [18] were superimposed on the Mn_2·Y-PPase·$(MnP_i)_2$ monomer 1 in O using the option lsq_exp [29]. The superpositions used sidechains, not Cα atoms of the residues. AP and Y-PPase were superimposed atom by atom the following sets of atoms were aligned (AP→Y-PPase): Asp369(Cγ)→Asp115(Cγ), Asp51(Cγ)→Asp152(Cγ), Asp327(Cγ)→Asp120(Cγ), Arg166(Cζ)→Lys56(Cε), Zn451→M1 and Zn450→M2, and the atoms superimpose with a rms deviation of 0.82 Å. The Mg^{2+} in AP and M4 in Y-PPase were not included in the superposition. Pol β was superimposed on Y-PPase the following sets of atoms were aligned (pol β→Y-PPase): Arg183(Cζ)→Arg78(Cζ), Asp190(Cγ)→Asp152(Cγ), DCT 338(Pα)→P2(P), Mg340→M1 and Mg339→M4. These atoms align with an rms deviation of 0.85 Å. The comparison of Klenow polymerase 3′,5′ exonuclease active site to Y-PPase was done by superimposing the following atoms (Klenex→Y-PPase): Asp355(Cγ)→Asp152(Cγ), Glu357(Cδ)→Asp115(Cγ), Asp424(Cγ)→Asp147(Cγ), Asp501(Cγ)→Asp120(Cγ), Wat933(O)→Wat1(O), dTMP(P)→P2(P), Zn930→M1, Zn931→M4. These atoms align with an rms deviation of 0.73 Å.

The secondary structure definitions and the Ramachandran plot are from PROCHECK [67] (Fig. 2 and Table 3) The Ramachandran plot shows that the mainchain geometry of the structures is very good. Only one residue, Asp66, in a long turn preceding strand β3, is in a generously allowed region, but the electron density shows that it is correctly modelled. Rms deviations of bond lengths and angles are from X-PLOR [60] (Table 3).

Accession numbers

The coordinates for the Mn_2·Y-PPase and Mn_2·Y-PPase·$(MnP_i)_2$ structures have been deposited with the Brookhaven Protein Databank, with accession numbers 1WGI and 1WGJ, respectively. The coordinates are on hold for one year. Supplementary material for the paper is available through the World Wide Web at http://www.btk.utu.fi/xray/PPase.

Acknowledgements

We thank Raija Andersén, Nils Kullberg and Dan-Johan Still for skilful technical assistance and Gerard Kleywegt for running his program SPASM for

us prior to its release. We also thank Andrew Leslie for the F_1-ATPase coordinates, Thomas Steitz for the Klenow polymerase coordinates and Donald Voet for the initial Y-PPase coordinates. Pirkko Heikinheimo was supported by Emil Aaltosen Säätiö and Jukka Lehtonen by the Suomen Akatemia Protein Structure and Function Graduate Program. This work was supported by Suomen Akatemia (grant numbers 3875 to RL, 1444 to AG and 4310 to AG and RL), by the NIH (DK13212 to BSC and TW00407 to AB and BSC), and by the Russian Foundation for Basic Research (94-04-12658-a to AB).

References

1. Chen, J., *et al.*, & Plateau, P. (1990). Pyrophosphatase is essential for growth of *Escherichia coli. J. Bacteriol.* **172**, 5686–5689.
2. Lundin, M., Baltscheffsky, H. & Ronne, H. (1991). Yeast PPA2 gene encodes a mitochondrial inorganic pyrophosphatase that is essential for mitochondrial function. *J. Biol. Chem.* **266**, 12168–12172.
3. Kornberg, A. (1962). On the metabolic significance of phosphorolytic and pyrophosphorolytic reactions. In *Horizons in Biochemistry*. (Kasha, M. & Pullman, B., eds), pp. 251–264. Academic Press, NY, USA.
4. Herschlag, D. & Jencks, W.P. (1990). Catalysis of the hydrolysis of phosphorylated pyridines by $Mg(OH)^+$ — a possible model for enzymatic phosphoryl transfer. *Biochemistry* **29**, 5172–5179.
5. Arutyunyan, É.G., *et al.*, & Hansen, G. (1981). X-ray structural investigation of inorganic pyrophosphatase from baker's yeast at 3 Å resolution. *Dokl. Akad. Nauk SSSR* **258**, 1481–1485.
6. Bernstein, F.C., *et al.*, & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structure. *J. Mol. Biol.* **112**, 535–542.
7. Chirgadze, N.Y., *et al.*, & Khene, V. (1991). Crystal structure of MnP complex of inorganic pyrophosphatase of yeast at resolution of 2.35 Å. *Sov. Phys. Cryst.* **36**, 128–132.
8. Harutyunyan, E.H., *et al.*, & Wilson, K.S. (1996). X-ray structure of yeast inorganic pyrophosphatase complexed with manganese and phosphate. *Eur. J. Biochem.* **239**, 220–228.
9. Kankare, J., Salminen, T., Lahti, R., Cooperman, B.S., Baykov, A.A. & Goldman, A. (1996). The structure of *E. coli* inorganic pyrophosphatase at 2.2 Å resolution. *Acta Cryst. D* **52**, 551–563.
10. Kankare, J., Salminen, T., Lahti, R., Cooperman, B.S., Baykov, A.A. & Goldman, A. (1996). Crystallographic identification of metal-binding sites in *E. coli* inorganic pyrophosphatase. *Biochemistry* **35**, 4670–4677.
11. Oganessyan, V.Y., *et al.*, & Harutyunyan, E.H. (1994). X-ray crystallographic studies of recombinant inorganic pyrophosphatase from *Escherichia coli. FEBS Lett.* **348**, 301–304.
12. Teplyakov, A., *et al.*, & Kuranova, I. (1994). Crystal structure of inorganic pyrophosphatase from *Thermus thermophilus*. *Protein Sci.* **3**, 1098–1107.
13. Arutyunyan, É.G., *et al.*, & Avaeva, S.M. (1996). The structure of *E. coli* inorganic pyrophosphatase and its complex with Mn^{2+} at 2.2 Å resolution. *Kristallographiya* **41**, 84–96.
14. Gonzalez, M.A., Webb, M.R., Welsh, K.M. & Cooperman, B.S. (1984). Evidence that catalysis by yeast inorganic pyrophosphatase proceeds by direct phosphoryl transfer to water and not via phosphoryl enzyme intermediate. *Biochemistry* **23**, 797–801.
15. Cooperman, B.S. (1982). The mechanism of action of yeast inorganic pyrophosphatase. *Methods Enzymol.* **87**, 526–548.
16. Baykov, A.A., *et al.*, & Cooperman, B.S. (1996). Catalysis by *Escherichia coli* inorganic pyrophosphatase: pH and Mg^{++} dependence. *Biochemistry* **35**, 4655–4661.
17. Lahti, R. (1983). Microbial inorganic pyrophosphates. *Microbiol. Rev.* **47**, 169–179.
18. Beese, L.S. & Steitz, T.A. (1991). Structural basis for the 3′–5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33.
19. Bujacz, G., *et al.*, & Skalka, A. (1996). The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* **4**, 89–96.
20. Patel, P., *et al.*, & Arnold, E. (1995). Insights into DNA polymerization mechanism from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry* **34**, 5351–6363.
21. Kim, E. & Wyckoff, H. (1991). Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J. Mol. Biol.* **218**, 449–464.
22. Kim, Y., Eom, S.H., Wang, J., Lee, D.-S., Suh, S.W. & Steitz, T.A. (1995). Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376**, 612–616.
23. Kankare, J., *et al.*, & Goldman, A. (1994). The structure of *E. coli* soluble inorganic pyrophosphatase at 2.7 Å resolution. *Protein Eng.* **7**, 823–830.
24. Lahti, R., Kolakowski, L.F., Heinonen, J., Vihinen, M., Pohjanoksa, K. & Cooperman, B.S. (1990). Conservation of functional residues between yeast and *E. coli* inorganic pyrophosphatases. *Biochem. Biophys. Acta* **1038**, 338–345.
25. Cooperman, B.S., Baykov, A.A. & Lahti, R. (1992). Evolutionary conservation of the active site of soluble inorganic pyrophosphatase. *Trends Biochem. Sci.* **17**, 262–266.
26. Baykov, A.A. & Shestakov, A.S. (1992). Two pathways of pyrophosphate hydrolysis and synthesis by yeast inorganic pyrophosphatase. *Eur. J. Biochem.* **206**, 463–470.
27. Heikinheimo, P., *et al.*, & Lahti, R. (1996). A site-directed mutagenesis study of *Saccharomyces cerevisiae* inorganic pyrophosphatase. Functional conservation of the active site of soluble inorganic pyrophosphatases. *Eur. J. Biochem.* **239**, 138–143.
28. Salminen, T., *et al.*, & Lahti, R. (1995). Structure and function analysis of *Escherichia coli* inorganic pyrophosphatase: is an hydroxide ion the key catalytic residue? *Biochemistry* **34**, 782–791.
29. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
30. Cooperman, B.S., Panackal, A., Springs, B. & Hamm, D.J. (1981). Divalent metal ion, inorganic phosphate, and inorganic phosphate analogue binding to yeast inorganic pyrophosphatase. *Biochemistry* **20**, 6051–6060.
31. Käpylä, J., Hyytiä, T., Lahti, R., Goldman, A., Baykov, A.A. & Cooperman, B.S. (1995). Effect of D97E substitution on the kinetic and thermodynamic properties of *Escherichia coli* inorganic pyrophosphatase. *Biochemistry* **34**, 792–800.
32. Heikinheimo, P., Salminen, T., Cooperman, B., Lahti, R. & Goldman, A. (1995). New crystal forms of *E. coli* and *S. cerevisiae* soluble inorganic pyrophosphatases. *Acta Cryst. D* **51**, 399–401.
33. Welsh, K.M., Armitage, I.M. & Cooperman, B.S. (1983). Yeast inorganic pyrophosphatase. Functional and $^{113}Cd^{2+}$ and ^{31}P nuclear magnetic resonance studies of the Cd^{2+}-enzyme. *Biochemistry* **22**, 1046–1054.
34. Springs, B., Welsh, K.M. & Cooperman, B.S. (1981). Thermodynamics, kinetics, and mechanism in yeast inorganic pyrophosphatase catalysis of inorganic pyrophosphate: inorganic phosphate equilibration. *Biochemistry* **20**, 6384–6391.
35. Cooperman, B.S. (1976). The role of divalent metal ions in phosphoryl and nucleotidyl transfer. In *Metal Ions in Biological Systems*. (Sigel, H., ed.), pp. 80–125, Marcel Dekker, NY, USA.
36. Benkovic, S.J. & Schray, K.J. (1973). Chemical basis of biological phosphoryl transfer. In *The Enzymes*. (Boyer, P., ed.), pp. 201–238, Academic Press, NY, USA.
37. Spiro, T.G. (1971). Phosphate transfer and its activation by metal ions: alkaline phosphatase. In *Inorganic Biochemistry*. (Eichhorn, G.L., ed.), pp. 578–581, Elsevier, Amsterdam.
38. Breinig, J.B. & Jones, M.M. (1963). The effect of coordination on the reactivity of aromatic ligands. VII. Specific ion effects on diazo coupling rates. *J. Org. Chem.* **28**, 852–854.
39. Knight, W.B., Fitts, S.W. & Dunaway-Mariano, D. (1981). Investigation of the catalytic mechanism of yeast inorganic pyrophosphatase. *Biochemistry* **20**, 4079–4086.
40. Hsu, C.H. & Cooperman, B.S. (1976). Metal ion catalysis of phosphoryl transfer via a ternary complex. The effects of changes in leaving group, metal ion, and attacking nucleophile. *J. Am. Chem. Soc.* **98**, 5657–5663.
41. Lindqvist, Y., Schneider, G. & Vihko, P. (1994). Crystal structures of rat acid phosphatase complexed with the transition-state analogs vanadate and molybdate. Implications for the reaction mechanism. *Eur. J. Biochem.* **221**, 139–142.
42. Sondek, J., Lambright, D.G., Noel, J.P., Hamm, H.E. & Sigler, P.B. (1994). GTPase mechanism of G proteins from the 1.7 Å crystal structure of transducin α-GDP–AlF^-_4. *Nature* **372**, 276–279.
43. Konsowitz, L. & Cooperman, B.S. (1976). Solvent isotope effect in inorganic pyrophosphatase catalyzed hydrolysis of inorganic pyrophosphate. *J. Am. Chem. Soc.* **98**, 1993–1995.
44. Volk, S.E., *et al.*, & Cooperman, B.S. (1996). The effect of E20D substitution in the active site of *Escherichia coli* inorganic pyrophosphatase on its quaternary structure and catalytic properties. *Biochemistry* **35**, 4662–4669.
45. Schulz, G.E. (1992). Binding of nucleotides by proteins. *Curr. Opin. Struct. Biol.*, **2**, 61–67.

46. Baltscheffsky, H., Alauddin, M., Falk, G. & Lundin, M. (1987). Similarities between soluble inorganic pyrophosphatase from yeast and some nucleotide-binding polypeptides. *Acta Chem. Scand. [B]* **41**, 106–107.
47. Abrahams, J.P., Leslie, A., Lutter, R. & Walker, J. (1994). Structure at 2.8 Å resolution of F_1-ATPase from bovine heart mitochondria. *Nature* **370**, 621–628.
48. Bairoch, A. (1991). PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res.* **19** (suppl), 2241–2245.
49. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
50. Pelletier, H., Sawaya, M., Kumar, A., Wilson, S. & Kraut, J. (1994). Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. *Science* **264**, 1891–1902.
51. Sawaya, M., Pelletier, H., Kumar, A., Wilson, S. & Kraut, J. (1994). Crystal structure of rat DNA polymerase β: evidence for a common polymerase mechanism. *Science* **264**, 1930–1935.
52. Date, T., Yamamoto, S., Tanihara, K., Nishimoto, Y. & Matsukage, A. (1991). Aspartic acid residues at positions 190 and 192 of rat DNA polymerase β are involved in primer binding. *Biochemistry* **30**, 5286–5292.
53. Welsh, K.M., Jacobyansky, A., Springs, B. & Cooperman, B.S. (1983). Catalytic specificity of yeast inorganic pyrophosphatase for magnesium ion as cofactor. An analysis of divalent metal ion and solvent isotope effects on enzyme function. *Biochemistry* **22**, 2243–2248.
54. Ollis, D.L., *et al.*, & Goldman, A. (1992). The α/β hydrolase fold. *Protein Eng.* **5**, 197–211.
55. Brannigan, J.A., *et al.*, & Murzin, A.G. (1995). A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* **378**, 416–419.
56. Baltscheffsky, H., Lundin, M., Luxemburg, C., Nyren, P. & Baltscheffsky, M. (1986). Inorganic pyrophosphatase and the molecular evolution of biological energy coupling. *Chemica Scripta B* **26**, 259–262.
57. Klemme, J.-H. (1976). Regulation of intracellular pyrophosphatase-activity and conservation of the phosphoanhydride-energy of inorganic pyrophosphate in microbial metabolism. *Z. Naturforsch. [C]* **31**, 544–550.
58. Rapoport, T.A., Höhne, W.E., Heitmann, P. & Rapoport, S. (1973). Binding if ligands to the inorganic pyrophosphatase of bakers' yeast. *Eur. J. Biochem.* **33**, 341–347.
59. Otwinowski, Z. (1993). DENZO: an oscillation data processing program for protein crystallography. Yale University Press, New Haven, CT, USA.
60. Brünger, A.T. (1987). *X-PLOR. Version 3.1: A System for X-Ray Crystallography and NMR*. Yale University Press, New Haven, CT, USA.
61. Priestle, J.P. (1994). Stereochemical dictionaries for protein structure refinement and model building. *Structure* **2**, 911–913.
62. Steigemann, W. (1974). Die entwicklung und anwendung von rechenverfahren und rechenprogrammen zur struckturanalyse von proteinen am beispiel des trypsin-trypsininhibitor komplexes, des freien inhibitors und der L-asparaginase. PhD thesis, Munich: Technical University, Germany.
63. Kleywegt, G.J. & Jones, T.A. (1994). Halloween ... masks and bones. In *From First Map to Final Model*. (Bailey, S., Hubbard, R. & Waller, D., eds), pp. 59–66, SERC Daresbury Laboratory, Warington, UK.
64. Kleywegt, G.J. (1995). Dictionaries for heteros. ESF/CCP4 *Newsletter* **31**, 45–50.
65. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.
66. Kleywegt, G.J. & Jones, T.A. (1993). Masks made easy. ESF/CCP4 *Newsletter* **28**, 56–59.
67. Laskowski, R.A., MacArthur, M.V., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
68. Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics* **8**, 52–56.
69. Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522–524.
70. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
71. Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Cryst. D* **50**, 869–873.
72. Wallace, A.C., Laskowski, R.A. & Thornton, J.M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.